Clustering-based Filtering to Detect Isolated and Intermittent Pulses in Radio Astronomy Data

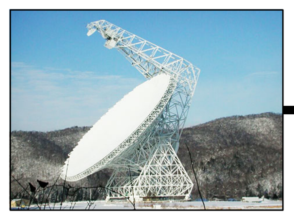
Kiri L. Wagstaff, Benyang Tang, Joseph Lazio, and Sarah Burke-Spolaor
Jet Propulsion Laboratory, California Institute of Technology
Contact: kiri.wagstaff@jpl.nasa.gov

Introduction

Many types of objects may emit radio pulses, including periodic pulsars, intermittent pulsars, evaporating primordial black holes, etc. The propagation of these signals through the interstellar medium creates a characteristic chirped (dispersed) signal. There exist well developed analysis methods for identifying periodic and single dispersed signals.

Problem: These searches generate far too many candidates for manual review to be practical.

Signals received by radio telescope



Analysis software (e.g., PRESTO) → Hundreds of thousands of candidates

Goal: Detect sources that emit isolated or intermittent pulses

Our solution: Use machine learning techniques to help filter and organize candidate detections for interpretation.

Example data set

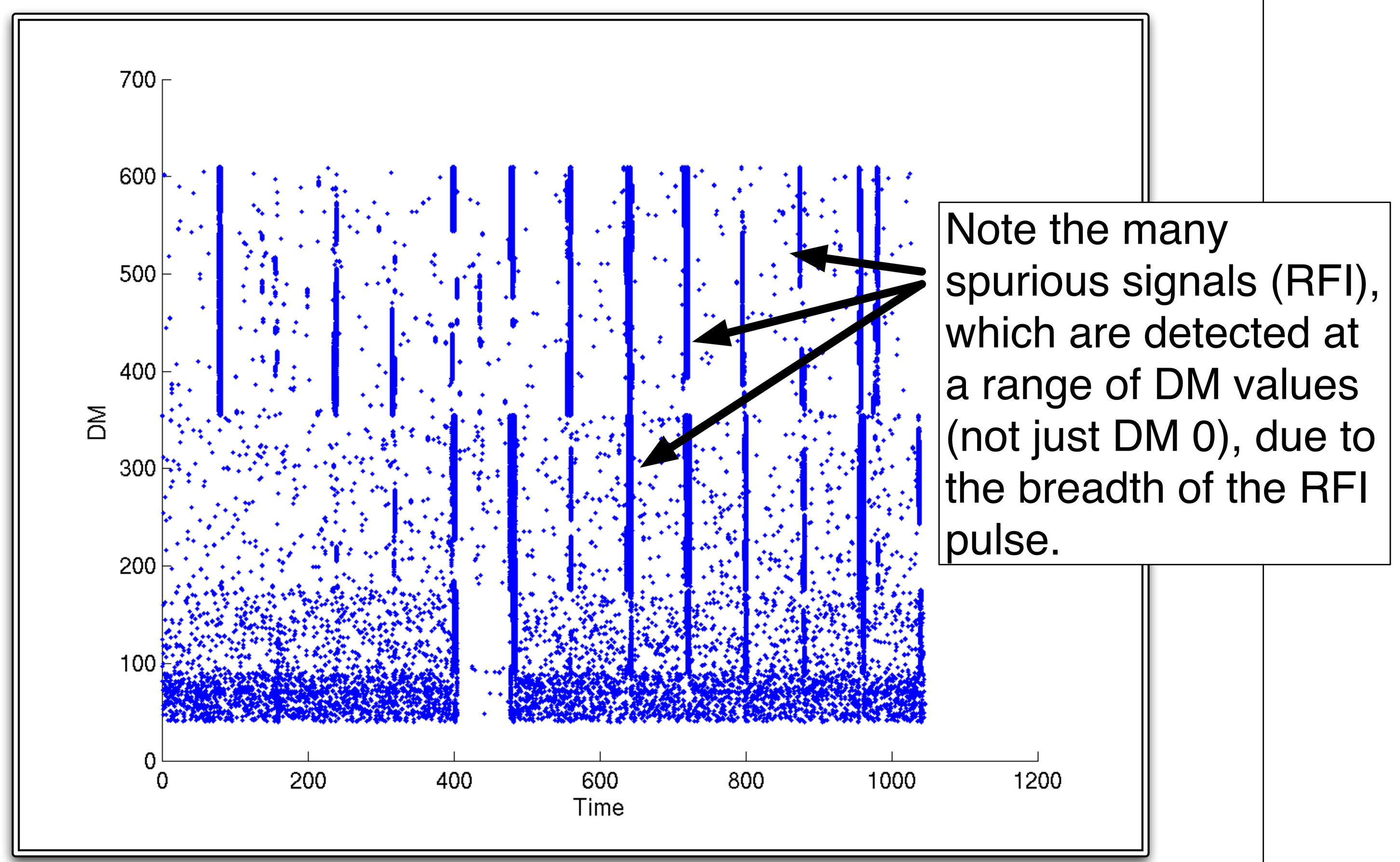


Fig. 1: Pulses detected in a test observation of pulsar PSR B1937+21 (DM ~71), acquired with the SPIGOT data acquisition system on the Green Bank Telescope. There are 227,568 detections from just 17.5 minutes of observing.

This is a particularly challenging case since the RFI itself is periodic!

We'd like to detect weaker, astronomical signals of scientific interest that are buried in this data set.

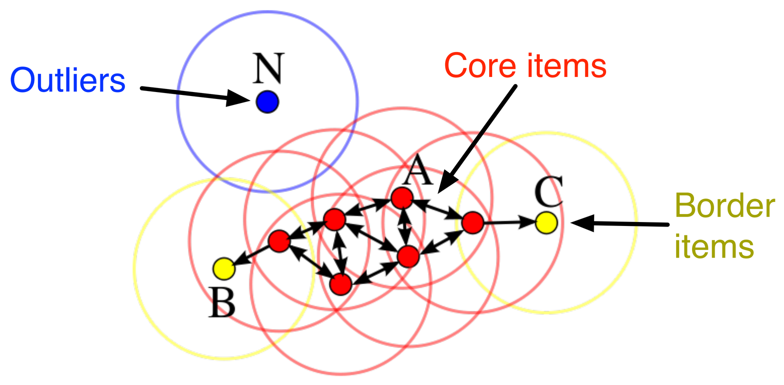
Method

Cluster the list of single pulse detections by DM, SNR, and pulse width using DBSCAN.

DBSCAN Clustering → Detections organized into a small number of groups, by similar properties

DBSCAN clustering [Ester et al., 1996]

- Density-based clustering
- Also identifies outliers



Benefits

- Efficient for large data sets
- No need to specify number of clusters in advance
- No model of expected cluster shape; instead model emerges from data
- Can find clusters that are not linearly separable

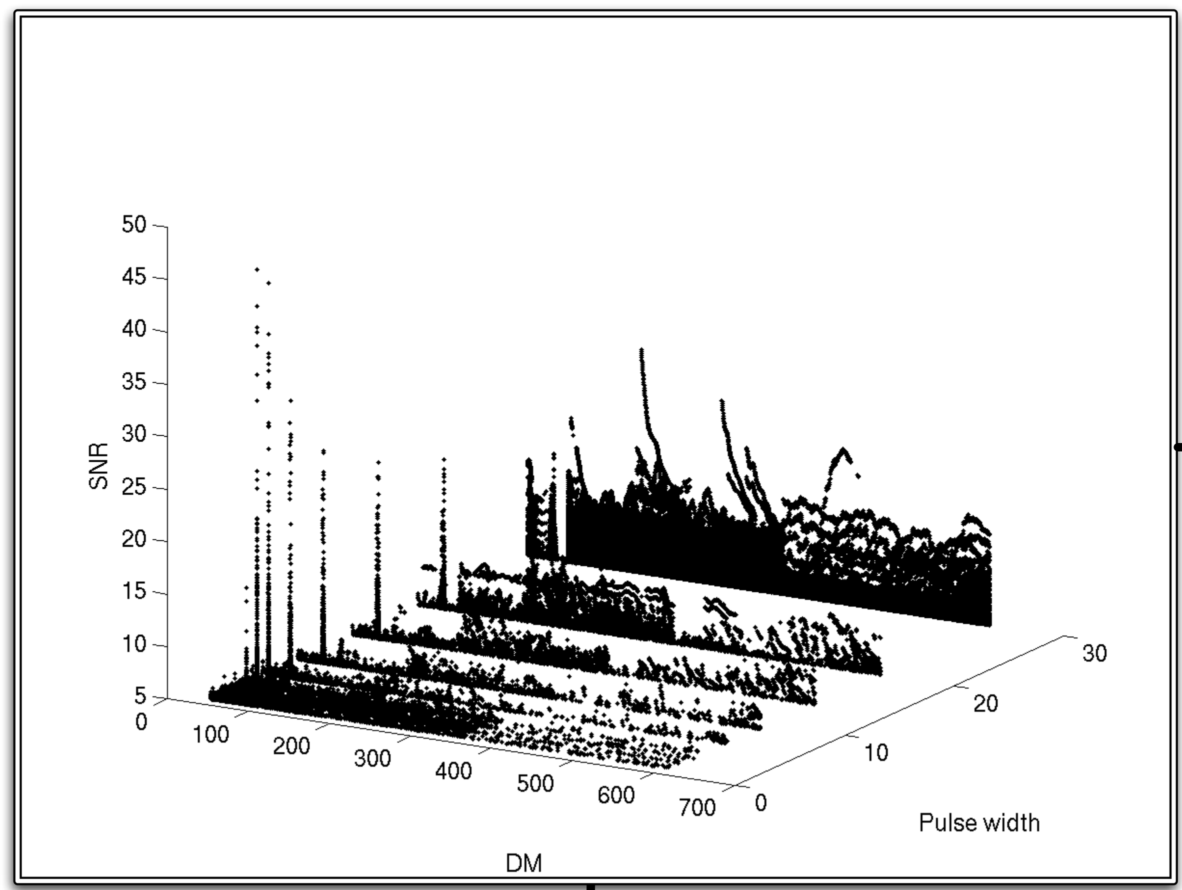


Fig. 2: PSR B1937+21 data in the feature space used by DBSCAN.

DBSCAN identified 7 clusters + outliers

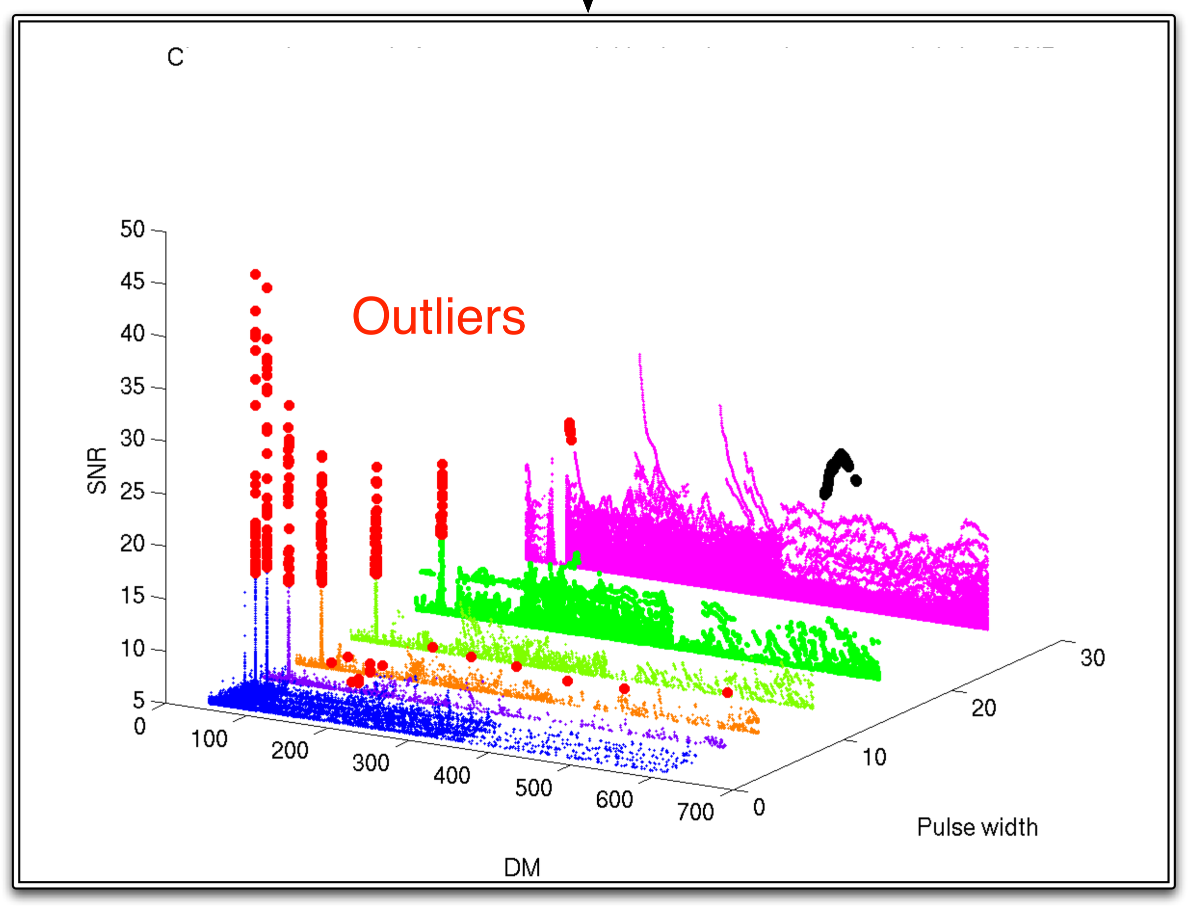


Fig. 3: PSR B1937+21 data in the feature space used by DBSCAN, color-coded by cluster assignment. Red points were deemed outliers by DBSCAN.

Example results

Cluster	# items	DM	SNR	Pulse width	Interpretation
Outliers	213	87	**19.0**	10	Pulsar
1	196,697	271	6.6	30	Likely RFI
2	12,534	264	6.8	20	Likely RFI
3	8743	145	5.5	2	Likely RFI
4	4612	280	5.9	14	Likely RFI
5	2818	235	5.7	9	Likely RFI
6	1842	173	5.6	6	Likely RFI
7	109	425	**18.2**	30	?

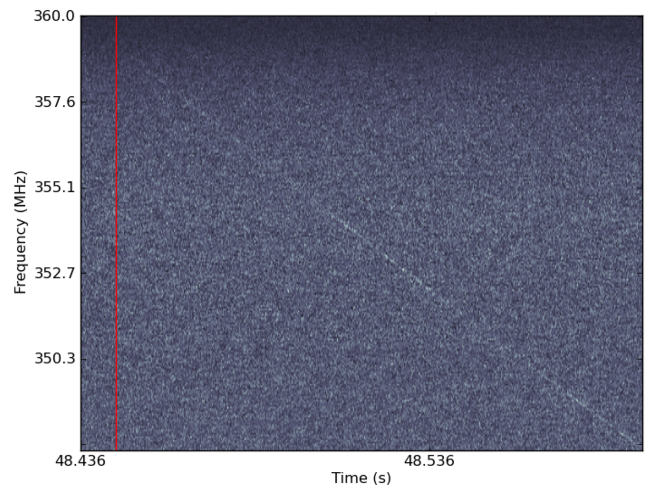


Fig. 4: Stacked visualization (to enhance SNR) of all detections assigned to the outlier cluster (pulses from PSR B1937+21). Only the top 250 frequency channels are shown.

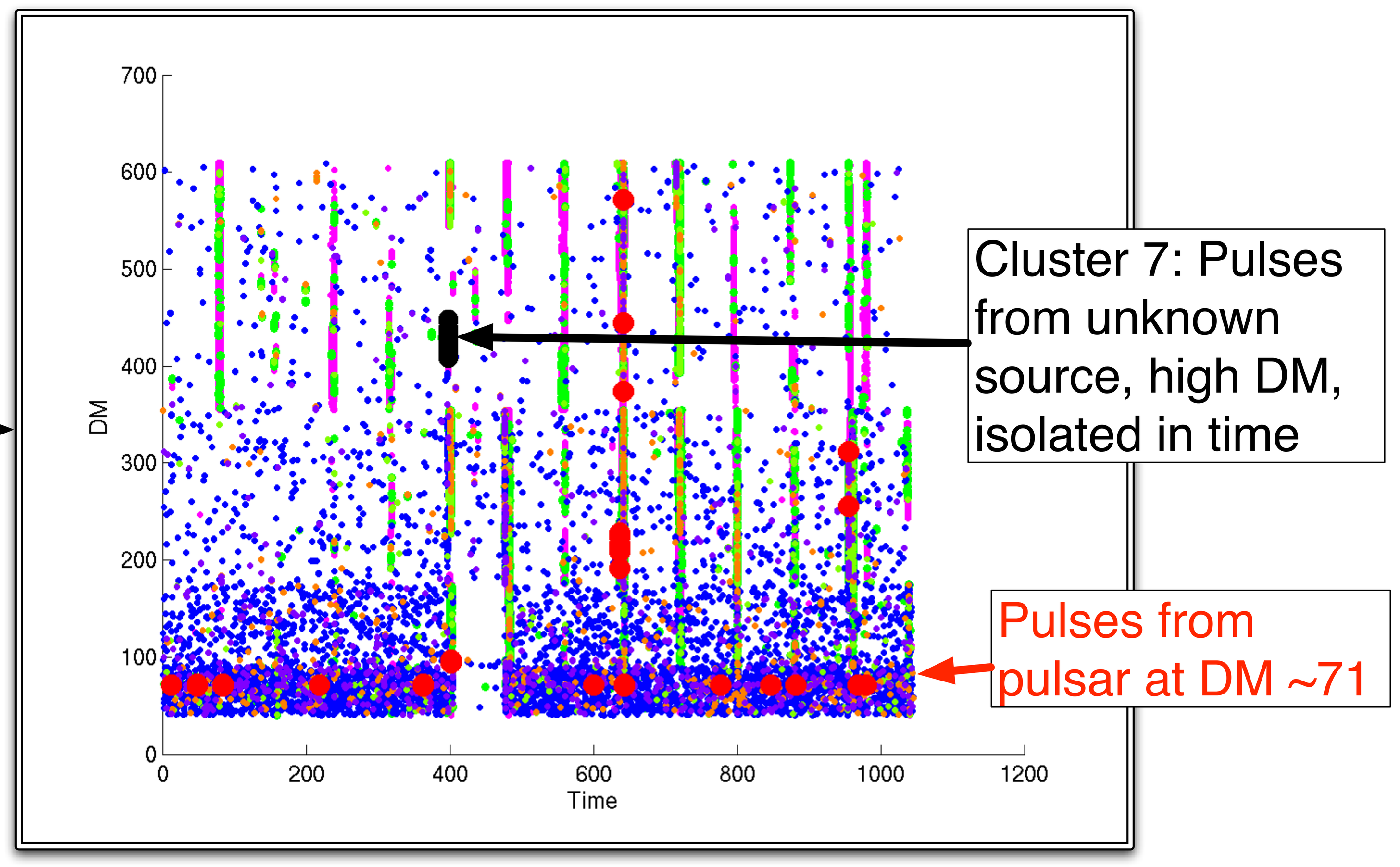


Fig. 5: PSR B1937+21 data plotted by DM and time, color-coded by DBSCAN cluster assignments.

Discussion

DBSCAN clustering can quickly organize large data sets to help identify interesting or unusual sub-groups (clusters) of data. For radio astronomy, it can aid in interpreting large single pulse detection lists. Significant time is saved in sifting through the data.

Clusters that are identified as likely RFI can be selectively masked and the remaining detections re-clustered to reveal finer distinctions.

Interested in trying this software on your data? Contact us. Code is available in Matlab, C, and Python.

This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA.